R E P O R T R E S U M E S

ED 012 009 AL 000 041

ON THE NOTION OF "EQUIVALENT SENTENCE STRUCTURE."

BY- FILLMORE, CHARLES J.
OHIO STATE UNIV., COLUMBUS, RESEARCH FOUNDATION

REPORT NUMBER POLA-11-3 PUB DATE JUN 65
EDRS PRICE MF-$0.09 HC-$0.56 14P.

DESCRIPTORS- *GRAMMAR, JAPANESE, CHINESE, ENGLISH, *TRANSFORMATION THEORY (LANGUAGE), *MACHINE TRANSLATION, *DEEP STRUCTURES, *CONTRASTIVE LINGUISTICS, PHRASE STRUCTURE, PROJECT ON LINGUISTIC ANALYSIS, COLUMBUS

IN THE DEVELOPMENT OF A PROGRAM FOR MECHANIZING TRANSLATION, SOME PROBLEMS WOULD REMAIN EVEN WHEN THE MORPHEMES AND GRAMMATICAL PROCESSES INVOLVED ARE REASONABLY EQUIVALENT IN THE TWO LANGUAGES. EQUIVALENT GRAMMATICAL RELATION HAS TO BE DEFINED BECAUSE IMPORTANT STRUCTURAL SIMILARITIES AMONG SENTENCES ARE FOUND MORE OFTEN IN THEIR DEEP STRUCTURE THAN IN THEIR SURFACE STRUCTURE. THUS, TWO SENTENCES ARE EQUIVALENT IF (1) THE UNDERLYING SIMPLE SENTENCES ARE TRANSLATIONS OF EACH OTHER, (2) THE RELATIONS BETWEEN THESE UNDERLYING SENTENCES AND A PORTION OF THE RESULTING SENTENCE ARE EQUIVALENT, AND (3) THE RESULTING SENTENCES ARE TRANSLATIONS OF EACH OTHER. HOWEVER, STRICT EQUIVALENCE AMONG GRAMMATICAL RELATIONS IS NOT TO BE FOUND. DISPARITIES IN SENTENCE STRUCTURE ARE ILLUSTRATED WITH SENTENCES FROM CHINESE, JAPANESE, AND ENGLISH, AND DEEP-SEATED SYSTEMATIC DIFFERENCES ARE FOUND. (IT)

ED012009

COMPLIMENTS OF THE PROJECT
ON LINGUISTIC ANALYSIS

RF Project 1685 - 4

THE OHIO STATE UNIVERSITY

RESEARCH FOUNDATION

1314 KINNEAR ROAD COLUMBUS, OHIO 43212

Project on Linguistic Analysis
Report No. 11

ENGLISH AND MANDARIN CHINESE:
DEFINITE AND INDEFINITE DETERMINERS
AND MODIFYING CLAUSE STRUCTURES
Sandra S. Annear
Division of Linguistics

INTERROGATION IN ENGLISH AND
MANDARIN CHINESE
Dale E. Elliott
Division of Linguistics

ON THE NOTION OF 'EQUIVALENT
SENTENCE STRUCTURE'
Charles J. Fillmore
Division of Linguistics

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

June 1965
NATIONAL SCIENCE FOUNDATION
Grant No. GN-174

ON THE NOTION OF 'EQUIVALENT SENTENCE STRUCTURE'

C. J. Fillmore

Important structural similarities among sentences, either within the same language or across languages, are much more to be found in their underlying or deep structure than in their surface structure. Sentences with very similar deep structure can in fact have strikingly unlike surface structure.

As an example of this, consider the English sentence

(1) What did you see?

and the equivalent sentence in Japanese

(2) Anata ga nani o mimasita ka?

Both of these sentences have as subject a word meaning 'second person'. Both have non-human interrogative words as objects. In both, the subject is contrasted with other possible subjects. Both have past tense verbs meaning "see". And both are questions. The only underlying differences involve the Japanese politeness and honorific systems: in Japanese the word chosen for "you" is a choice among several possible words, relating to differences in levels of politeness or formality; and the word for "see" contains a 'politeness' morpheme mas.

Superficially, the sentences are of very different form. The category 'question' is exhibited in the English sentence as the morpheme order 'past + you' ("did you") as opposed to 'you + past'; in Japanese it is revealed by the presence of the sentence final particle ka. The interrogative word remains in the normal object position in Japanese, but in English it is obligatorily placed at the front of the sentence. In the English sentence 'past tense' is shown in the word "did", in the Japanese by the suffix ta. The 'contrast' in the subject is indicated by the preposition ga in Japanese, by 'contrastive stress' in English. Superficially, the sentences could hardly be less similar.

Given observations such as these, it is clear that the most reasonable conception of the process of inter-lingual translation would specify the following steps:

(I) analysis of the source language sentences to recover their deep structures;

(II) mapping of these structures into deep structures in

AL 000 041

the target language;

(III) conversion of the new deep structures into their surface structures through the application of transformations in the target language.

Step III is provided by an adequate grammar of the target language. Step I is possible to the degree that analysis algorithms work. The whole process is feasible at all only if Step II -- the mapping of the deep structures of one language into those of another -- is relatively simple and straightforward. The purpose of the present paper is to show some of the ways in which this is not the case.

There are many well understood difficulties in the development of a program for mechanizing translation. We are familiar with the classic examples of words in one language which do not have equivalents in some other language. We know, too, that semantically relevant grammatical categories may be obligatory in one language, optional in another, resulting inevitably (where readability is preserved) in a certain amount of information loss and information gain. All of these difficulties, however, are of equal seriousness under any translation scheme. For my purposes I shall disregard such problems and concentrate on those cases where both the morphemes and the grammatical processes involved are reasonably equivalent in the two languages. There will still be problems enough.

We may begin by developing the notion of 'equivalent grammatical relation'. First of all, we assume that every grammar defines 'relatedness' among sentences. Relatedness between two sentences may be stated in such ways as that one sentence contains a morpheme which the other sentence lacks, or one sentence has undergone a transformation which the other has not, or one sentence contains the other as one of its parts.

Some sentence relations have two terms, as between the active and the passive forms of the sentence. We may represent such a relation diagrammatically as follows:

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

(S_1)

P

(S_2)

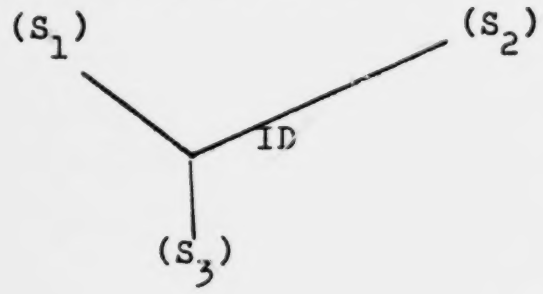
where 'P' stands for 'passive'. This relation holds where S_1 is

(3) The dog bit the boy.

and where S_2 is

(4) The boy was bitten by the dog.

Other sentence relations are multiple, as seen in the case of a sentence S_3 which is related to (consists of) a sentence S_2 which is embedded in a sentence frame S_1. A three term relation of this type might be exemplified by the 'indirect discourse' relation of English, schematized in the following diagram --



where if S_1 is

(5) John said -- .

and S_2 is

(6) Mary is sick.

S_3 must be

(7) John said that Mary was sick.

Grammatical relations in two different languages can be called 'equivalent' whenever corresponding terms of the relations are translations of each other. In this sense the English and Japanese 'assertion/question' relations are equivalent because, in the majority of cases, where these relations hold and the assertions are translations of each other, the questions are also translations of each other.

Similarly the English process for forming relative clauses (i.e., the relation between a sentence and each of its relative clause forms)

is equivalent to that in Japanese. The Japanese sentence

(8) Ano hito ga tegami o kaita.

translates the English sentence

(9) That person wrote a letter.

In English a relative clause modifying "letter" can be formed -- roughly speaking -- by deleting "a letter" from the sentence, putting "which" in front of it, and placing the resulting sequence to the right of the noun "letter" in some sentence. The resulting noun phrase is

(10) the letter which that person wrote

The Japanese clause modifier is formed by deleting tegami o from the original sentence and placing what remains to the left of the word tegami in some sentence, giving the noun phrase

(11) ano hito ga kaita tegami

The two original sentences are translations of each other, and the resulting noun phrases are translations of each other. Since this would hold, generally, no matter what sentences we started out with, we may say that the English and Japanese processes for forming relative clauses are equivalent.

Now the notion of 'equivalent sentences' in two languages can be constructed as follows: two sentences are 'equivalent' if (1) the underlying simple sentences are translations of each other, (2) the relations between these underlying sentences and (a portion of) the resulting sentence are equivalent as defined above, and (3) the resulting sentences are translations of each other.

The definition of equivalent grammatical relations contained qualifying phrases such as "generally" or "in the majority of cases". No 'statistical' sense is intended by these qualifications. It seems to be true, however, that for certain grammatical processes which are equivalent in simple cases, there will be combinations of these processes which are possible in one language and impossible in another. Strict equivalence among grammatical relations, in other words, is not to be found. Disparities between linguistic systems exist which make it

impossible to identify equivalences between sentences even in some cases where the lexical elements and the grammatical relations involved are 'equivalent' to each other. There are cases where translation is possible, but where the two sentences are constructed along non-analogous principles, and there are also cases where sentences in one language cannot be translated into the other language at all.

We shall examine a few cases of structural non-equivalences involving Mandarin Chinese, English, and Japanese, but first it may be necessary to remind ourselves that certain apparent differences between linguistic systems are not really differences of the kind that need concern us. For example, at first we might think it important that while English permits the conjoining of two or more verbs before a single object, Chinese does not. That is, in English we can say

(12) I read and enjoyed your book.

but the Chinese sentence would have to be the equivalent of

(13) I read your book and enjoyed your book.

In their deep structures, however, the two sentences are equivalent, amounting to

(14) I read your book and I enjoyed your book.

In the collapsed form of this conjunction of sentences, both languages permit deletion of one of the repeated subjects; only English permits the deletion of one of the repeated objects. The difference is one of transformational detail only.

A similar argument will hold in connection with constructions associated with the English word "respectively". It may be true that in many languages there is no 'exact translation' of the English sentence

(15) John and Mary dislike peas and carrots respectively.

All languages, however, permit conjunction of sentences, including sentences which are partly similar. English merely happens to have a special rule which permits the collapsing of conjunctions of sentences that differ from each other in exactly two points, a rule which marks the result of this collapsing with the word "respectively". The difference

is one of surface structure only.

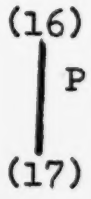
Structural properties of languages that are more apt to introduce difficulties can be illustrated by various combinations of 'causative' and 'passive' constructions in English. If we start out with the sentence

(16) John wrote a letter.

we know that its passive is

(17) A letter was written by John.

The relation between these two sentences may be symbolized as follows:



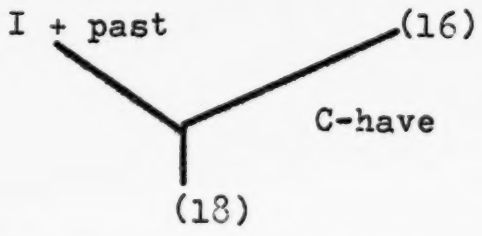
The causative, with "I" as instigator, can be formed on sentence (16) in either of two ways, giving us either

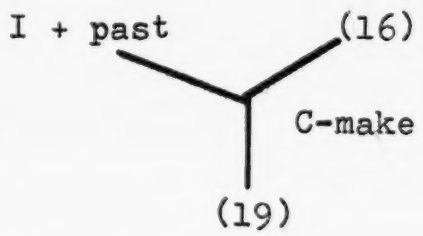
(18) I had John write a letter.

or

(19) I made John write a letter.

That is, we have to recognize at least two causative processes in English; it will be observed that in many cases the translation into English of a causative sentence in some other language may use either of these two processes. If the original causative is a 'coercive causative', the English causative "make" is obligatory; otherwise the causative with "have".





If the second term of a causative construction is to be a passive sentence such as (18), only the "have" causative is possible. That is

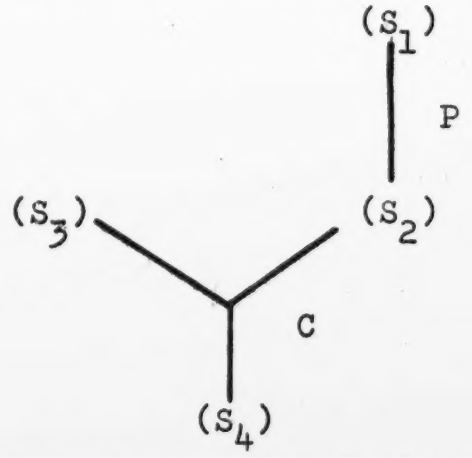
(20) I had a letter written by John.

is a grammatical sentence in English, but

(21) *I made a letter written by John.

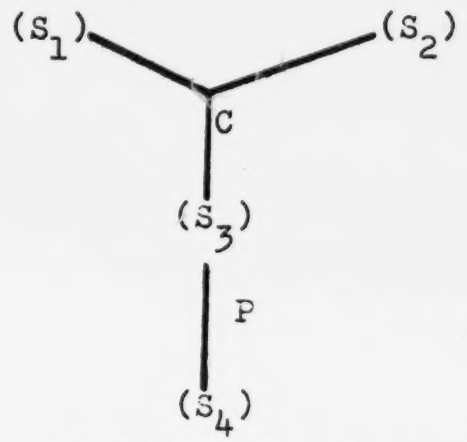
(interpreted as a causative) is not. What this means is that a sentence

whose underlying structure is of the following form:



cannot be translated by either of the English 'causative' devices considered so far if the original is a 'coercive causative'. It will be necessary to choose another 'coercive causative' verb such as "force".

That, of course, can be done. But what if the sentence is not the causative of a passive but the passive of a causative? That is, what if it had the following form?



As it happens, sentence 19 ("I made John write a letter.") may be made passive, but sentence 18 ("I had John write a letter.") may not. That is

(22) John was made to write a letter.

is grammatical, but

(23) *John was had to write a letter.

is not. What this means is that the passive form of a 'non-coercive causative' cannot be expressed in English.

The subtleties here, then, are that of the two general devices for forming causatives in English, only one can operate on a passive sentence, only the other can itself be made passive. If 'passive' is

not a meaning-changing transformation, then we are claiming not that translation is impossible in these cases, but that the full underlying structure of the source sentence must be perceived before the permitted combination of 'equivalent' processes in the target language can be selected.

Turning now to a different order of difficulties, we may consider various uses of interrogative morphemes. The conditions under which it is possible to interrogate constituents of sentences differ quite arbitrarily from language to language.

Differences between Japanese and English are particularly striking. In Japanese

(24) Kayoobi no sinbun desu.

means

(25) It's Tuesday's newspaper.

where ka ("fire") + yoobi ("day-of-the-week") means "Tuesday". Replacing ka by sui ("water") we get

(26) Suiyoobi no sinbun desu.

meaning

(27) It's Wednesday's newspaper.

Now it happens that in Japanese the morpheme identifying the day of the week may be interrogated, so that it is possible to ask the question

(28) Naniyoobi no sinbun desu ka?

where the English translation would have to be

(29) *What day's newspaper is it?

and that is not English.

In most simple cases, 'constituent interrogation' is equivalent in Japanese and English, and the same is true of 'relative clauses'. Certain combinations of these two processes, however, do not match in the two languages. Consider the Japanese sentence

(30) Kare ga kaita tegami o yomimasita.

which may be analytically rendered as "He" + 'subject' + "wrote" +

"letter" + 'object' + "(somebody) read", translated as

(31) (You) read the letter which he wrote.

In both languages it is possible to interrogate the noun which is modified by the relative clause giving us

(32) Kare ga kaita nani o yomimasita ka?

and

(33) What did you read that he wrote?

which are translations of each other. Only in Japanese, however, is it possible to interrogate a constituent of the relative clause itself. That is, the perfectly acceptable Japanese sentence

(34) Dare ga kaita tegami o yomimasita ka?

-- in which the subject of the relative clause has been interrogated -- cannot be directly translated into English, since

(35) *Who did you read a letter which wrote?

is ungrammatical. The sentence

(36) You read a letter which <u>who</u> wrote?

is acceptable English, but it is not -- as the Japanese sentence is -- a normal information question. It is rather a request to repeat something which one has not heard or does not believe.

The limitations on interrogation in connection with nouns modified by relative clauses separate English, Chinese, and Japanese from each other in interesting ways. Although there are numerous specific conditions limiting interrogation in each of these languages, their differences in this respect can be roughly stated as follows:

<u>Japanese</u>

permits either the noun modified by a relative clause or a constituent of the relative clause to be interrogated

<u>English</u>

permits the modified noun to be interrogated, but nothing from the relative clause

<u>Chinese</u>

permits constituents of the relative clause to be interro-

gated, but not the modified noun itself.

	Possible to interrogate nouns modified by relative clause	Possible to interrogate constituents of relative clause
English	+	-
Chinese	-	+
Japanese	+	+

In the English sentence

(37) You know someone who lives in this neighborhood.

It is possible, by interrogating the "someone", to form the question

(38) Who do you know who lives in this neighborhood?

This sentence may be translated 'directly' (i.e., it has an 'equivalent') in Japanese, as

(39) Kono hen̄ ni sun̄de īru dare o sitte īmasu ka?

but it cannot be directly translated into Chinese.

On the other hand, the Japanese sentence

(40) Kore wa dare ga kaita hon̄ desu ka?

translates into Chinese as

(41) Jèige shì shéi xiě de shū.

but it has no translation into English.

(42) *Who is this a book which wrote?

To take another syntactic feature related to relative clauses that is shared by Japanese and Chinese, we notice that in these languages a noun in a subordinate clause may be the 'shared noun' in the formation of a relative clause. The grammar of English does not allow a sentence like

(43) I cry every time I see that movie.

to be the basis of a relative clause modifying "movie", as in

(44) *It's a movie which I cry every time I see.

because the noun phrase "that movie" is contained in a subordinate clause. In both Japanese and Chinese, however, constructions of this type are perfectly normal. The Japanese sentence

(45) Sono eiga o miru tabi ni namida ga deru.

is a translation of sentence 43, where sono eiga ("that movie") is part of the subordinate clause sono eiga o miru tabi ni ("every time I see that movie"). The Japanese sentence, however, can be formed into a relative clause modifying eiga ("movie"), as in

(46) Miru tabi ni namida ga deru eiga desu.

which would have to be rendered in English as the ungrammatical sentence 44. The Chinese equivalent of sentence 46 is

(47) Zhè jiù shì wǒ yí kàn jiù kū de diànyǐng.

These sentences can be rendered in English, of course, but only by recasting them in such a way that the shared noun is in the main clause, not the subordinate clause, of the sentence which forms the relative clause. Instead of 44 we need a sentence like

(48) That movie makes me cry every time I see it.

From 48 we may form the relative clause found in

(49) It's a movie which makes me cry every time I see it.

I have not been claiming with these random remarks that there are ideas or propositions which can be expressed in some languages but not in others. That may be true, but it is not what I have been claiming. What I am saying is that beliefs about language universals which underlie the concept of mechanical translation necessarily include certain assumptions about the magnitude of the problem I have been discussing here. One of the tasks of the Ohio State University Project on Linguistic Analysis is to estimate this magnitude by cataloging disparities in sentence structures among the three languages Chinese, Japanese, and English. So far, our impression is that deep-seated systematic differences abound.